Assessing the Performance of Space Weather Models Using Metrics

K. A. Keller, M. Hesse, A. Narock
L. Rastaetter, M. M. Kuznetsova,
NASA Goddard Space Flight Center

Need for Metrics

- Create objective measure of current capabilities both for scientific and operational needs.
- Measure the improvement of model capabilities over time.
- Provide an objective comparison between models with comparable output.

Metrics which lead to scores near unity now are useless!

Elements of a Metric

- An output parameter from a model.
  - Example: Density or velocity at a satellite position
- A satellite or ground-based measurement that can be used for comparison.
  - Example: Plasma data from ACE
- A quantifiable norm that assesses the difference between the parameter from the model and the measurement.

Current Metrics

- Heliospheric metric using density and velocity from ACE.
- Ionospheric metric using data from ground magnetometer chains.
- Inner magnetosphere metric using particle fluxes at geosynchronous orbits from Los Alamos National Laboratory satellite data.

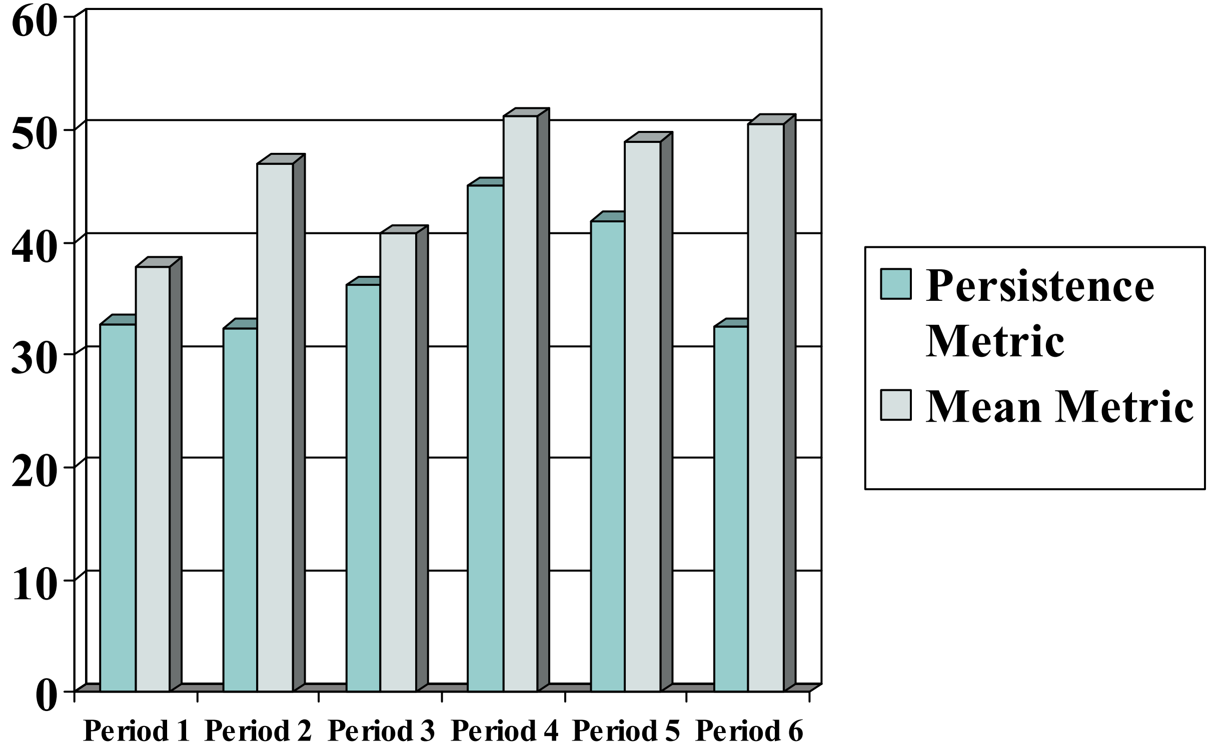
Heliosphere Metric

- Data
  - ACE velocity and density average every 6 hours for 27 days.
- Model
  - Heliospheric Tomography Model developed by B. Jackson and P. Hick. This model gives output every 6 hours for 27 days.
- Metric
  - A model is scored using $D_i = \text{sqrt}\,(\Sigma|\Delta H_{model} - \Delta H_{data}|^2/\text{npts})$.
  - A skill score is computed by

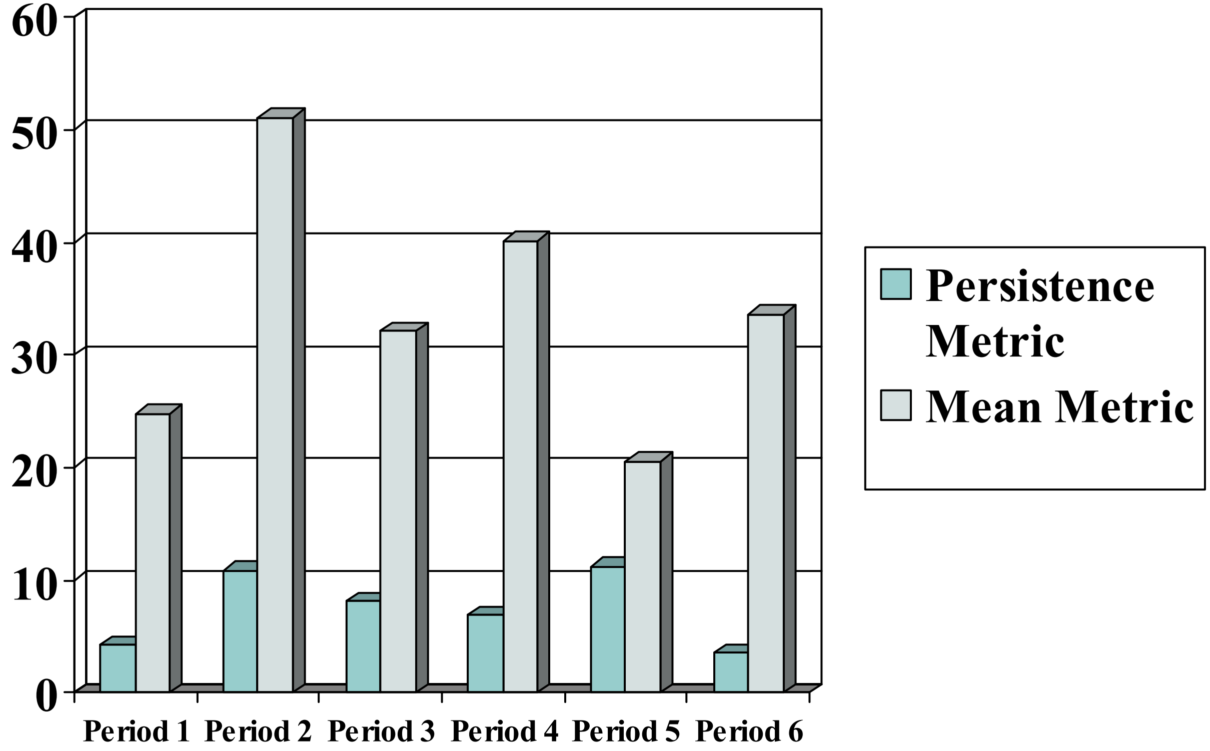
$$M_i = 1 - D_i / D_s$$

    where D_s is for the standard model. In this case, two standard models were used. One standard is a persistence metric which uses the previous measurement as the prediction for the current time step. The second standard is the mean for the entire Carrington rotation.
  - The score is then scaled so that the score is between 0 and 100 by the following transformation $S_i = 50 * (2^{M_i})$

Scores for Density



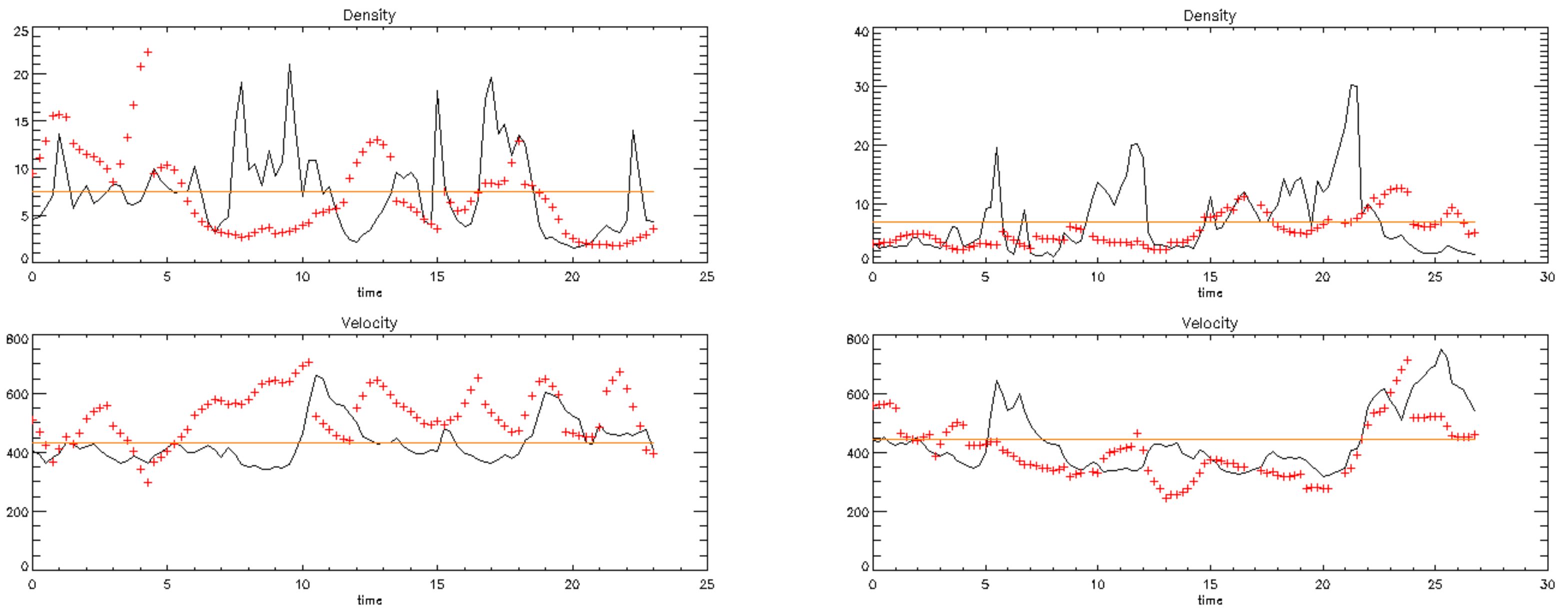
Scores for Velocity



Model and Data Comparison

Density

Velocity



Time (Days)

Period 1

Time (Days)

Period 2

Black: Data Red: Model Orange: Mean

We thank the ACE SWEPAM instrument team and the ACE Science Center for providing the ACE data.

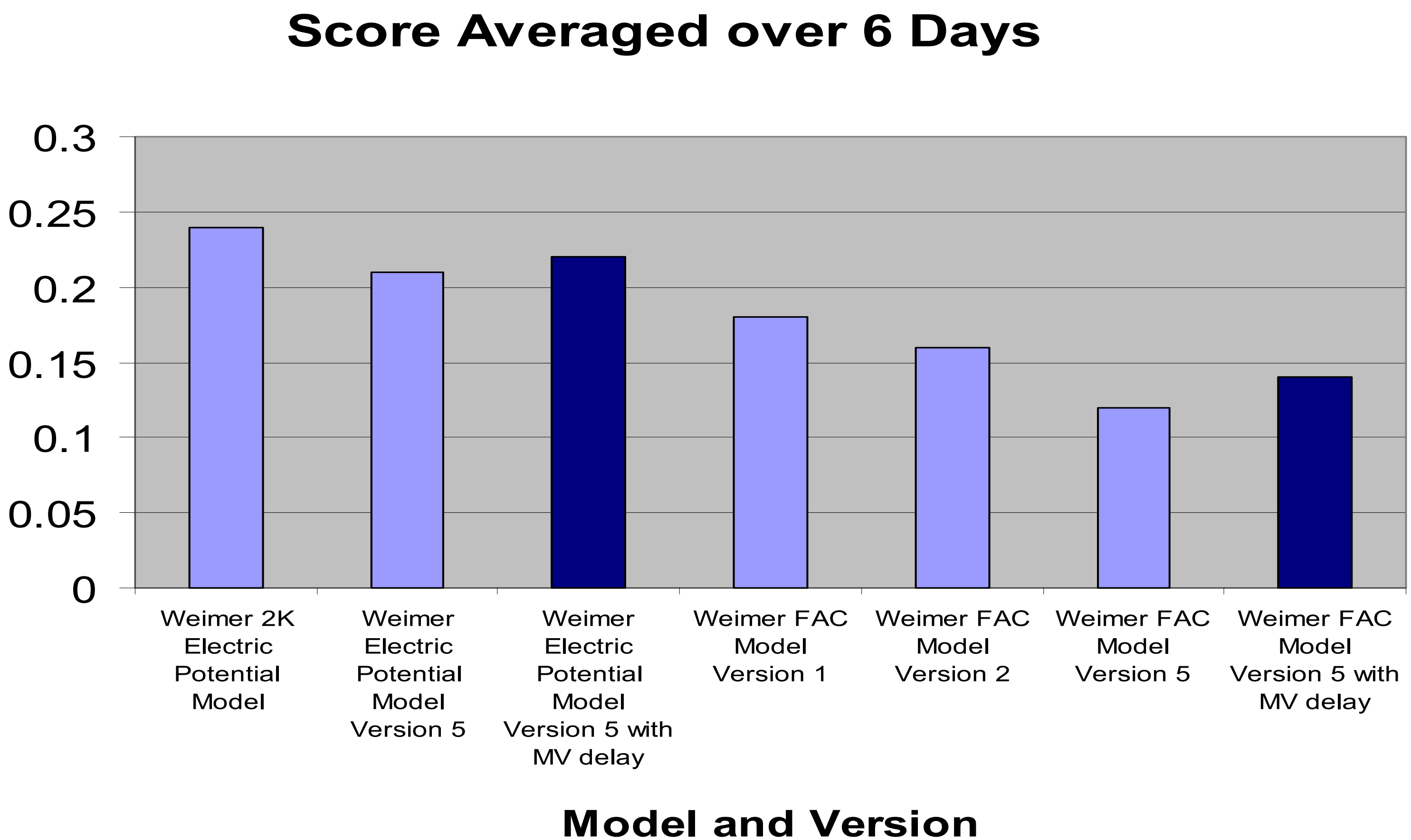
Ionospheric metric

- Data
  - Ground magnetic perturbations measured at 10 stations in the Greenland chain using the H component of the data.
- Models
  - Weimer electric potential model (2 different versions).
  - Weimer field-aligned current model (3 different versions).
- Skill score
  - An individual model is scored $D_i = \Sigma|\Delta H_{model} - \Delta H_{data}|/npts$.
  - A skill score is computed for each ground station by

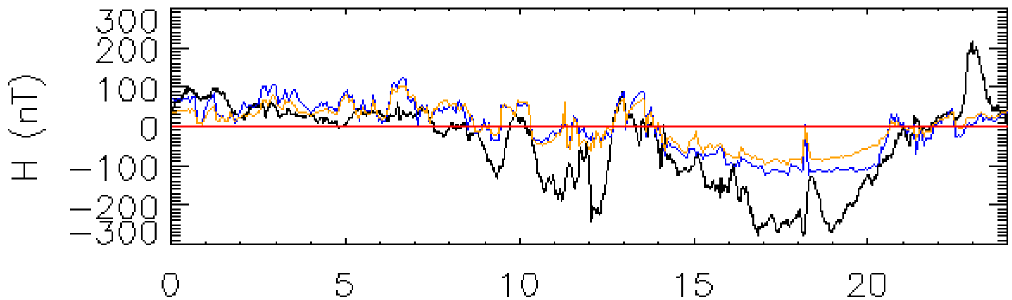
$$M_i = 1 - D_i / D_s$$

where D_s is for the standard model. In this case, the standard model is $\Delta H_{standard} \equiv 0$.

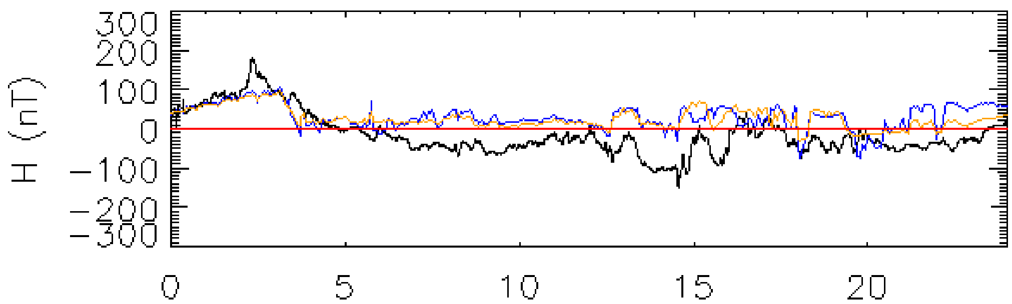
Results for Weimer Models (averaged over 10 stations) for H component.



Comparison of Model Results to Data



Time (hours)



Black: Data from ground magnetometers

Orange: Model results from Weimer 2k Electric Potential Model

Blue: Model results from Weimer Electric Potential Model Version 5

Magnetometer data was provided by the Danish Meteorological Institute (Dr. Jurgen Watermann, Project Scientist)

Inner Magnetospheric Metric

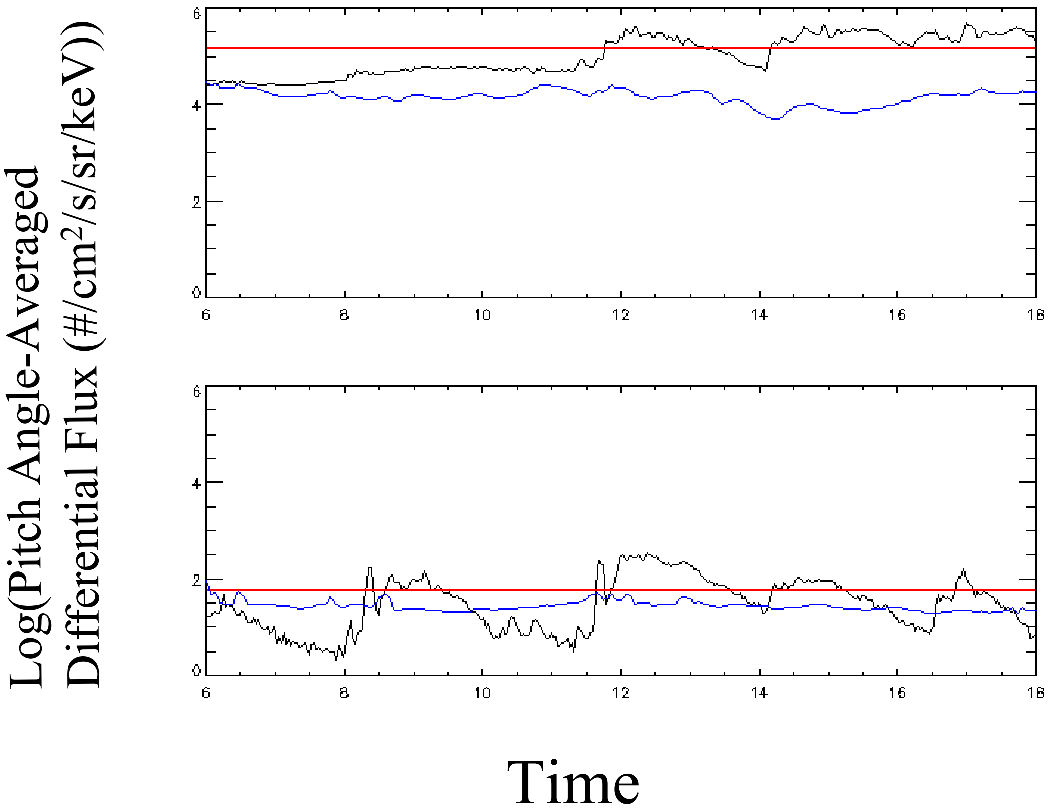
- Data
  - Proton fluxes from LANL geosynchronous satellites
- Model
  - Fok Ring Current model driven by a MHD model
- Skill Score using the Root Mean Square Deviation
  - Calculate mean square error

    $\text{RMS\_deviation} = \text{sqrt}\{\sum(\log_{10}[\text{predicted}/\text{observed}])^2/\text{npts}\}$
  - Calculate variance of observations

    $\text{STD\_deviation} = \text{sqrt}\{\sum(\log_{10}[\text{observed}/\text{mean}])^2/\text{npts}\}$
  - Skill score

    Skill score = 1- RMS_deviation/STD_deviation

Sample of Ring Current Skill Scores Sawtooth



Black is LANL data. Blue is the model results.

Energy Band (keV)	Root Mean Square
50-75	-.995
250-400	.232

Geosynchronous proton flux data was provided by the Energetic Particle team at Los Alamos National Laboratory, Richard Belian (PI).

Future Plans

- Inner magnetosphere
  - Extend ring current study to several events
  - Perform similar analysis for Fok Radiation Belt Model
- Global magnetosphere models
  - Comparison with GOES magnetic field data
- Heliosphere
  - Extend metric to new models
- Solar
  - Explore metric options based on limited data
- Ionosphere
  - Total Electron Content

Summary

- A persistence model is better at predicting velocity and density at the ACE satellite than the Heliospheric Tomography model. More scintillation data may improve the results of the model.
- The ground magnetic perturbations is a first attempt at a repeatable metric to compare different versions of a model.
- Fine tuning of metrics is required in collaboration with the operational agencies and researchers.
- These metrics are first steps at establishing a baseline for future versions and models.